The Collaboratory for Multi-scale Chemical Science (CMCS)

Larry Rahn-SNL*, Christine Yang, Carmen Pancerella, Wendy Koegler, David Leahy, Michael Lee, Renata McCoy, Theresa Windus-PNL*, James D. Myers, Karen Schuchardt, Brett Didier, Eric Stephan, Carina Lansing, Al Wagner-ANL*, Branko Ruscic, Michael Minkoff, Sandra Bittner, Gregor von Laszewski, Reinhardt Pinzon, Sandeep Nijsure, Kaizar Amin, Baoshan Wang, William Pitz-LLNL*, David Montoya-LANL*, Bill Barber, Lili Xu, Yen-Ling Ho, Thomas C. Allison-NIST*, William H. Green, Jr. -MIT*, Michael Frenklach-UCB*
* denotes Institutional Point of Contact

Summary

The goal of the CMCS is to enhance chemical science research by breaking down the barriers to rapid sharing of validated information and by opening new paradigms for multi-scale science. To accomplish this CMCS is developing an adaptive informatics infrastructure and demonstrating proof-of-concept by publicly deploying an integrated set of key collaboration tools and chemistry-specific applications, data resources, and services. We have demonstrated these capabilities at SC2002, implemented Release 1 of the software, produced use-cases illuminating the central aspects of the project, and are supporting two new user groups.

The Collaboratory for Multi-scale Chemical Science (CMCS) is developing new infrastructure and data-sharing concepts and piloting them among a multi-disciplinary team of chemical scientists working to advance combustion science. The chemical scientists focus on different physical scales in the combustion problem and are geographically distributed. The physical scales range from the electronic structure of atoms and molecules to direct simulations of turbulent combustion phenomena that occur in engines or industrial processes. The proof-of-concept capabilities that CMCS is building are also applicable to many other research areas. These capabilities rely on a capable production network infrastructure that provides user-friendly data and security services, supported collaboration tools, and support for compute intensive applications.

A portal serves as the web interface for the adaptable informatics infrastructure being developed by the CMCS team. The data infrastructure takes advantage of a variety of standards and open-source information technologies to provide an unprecedented ability to share data, data pedigree, and project information within groups and across communities. The portal, which can easily be enhanced and customized through the inclusion of new 'portlets,' includes real-time collaboration capabilities, search and notification tools, and a pedigree browser. To support the chemistry community, the CMCS team has integrated a variety of

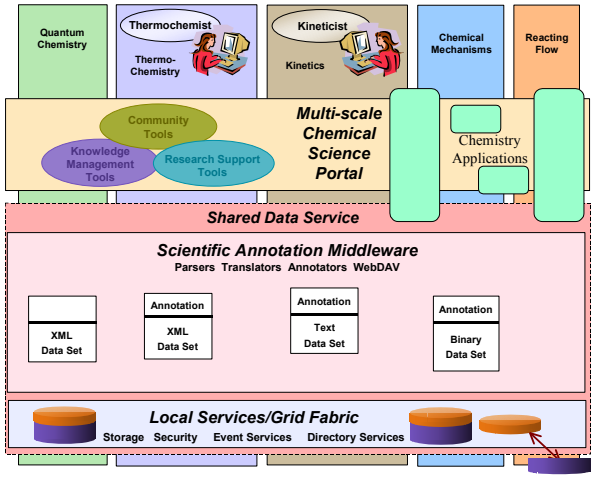

Figure 1. A diagram representing the major conceptual elements of the CMCS informatics infrastructure

* (925) 294-2091, rahn@sandia.gov

powerful chemistry applications, data viewers, and data translators. Figure 1 shows how CMCS users interact primarily with the top layer, the CMCS portal and chemistry applications. The applications can appear within the portal or provide their own user interfaces that interact directly with the underlying metadata/data and other CMCS services. The portal provides an array of functionality to support group and community processes, with an emphasis on simplifying the discovery and use of data. The shared data service, shown as the second layer, provides configurable capabilities for automating the generation of metadata, translating data between standard formats, and federating multiple data stores. At the lowest layer, the portal can take advantage of existing distributed services for security, event management, and data storage.

Data pedigree is at the heart of the CMCS project and is a key technology enabling new approaches to science. It allows researchers to categorize and trace scientific data across disciplines and scales and to identify the ultimate origin of scientific data. We have developed a DAV-aware pedigree browser (see lower screen shot in Fig. 2), which can easily display pedigree data (as well as annotations) and allows users to search, browse, and retrieve a data set's pedigree. Pedigree data and metadata are associated with CMCS resources by placing this information into the DAV properties of the data file in the CMCS data store. Pedigree data may also be an active link to a different, but associated, data resource.

Multiple science areas have made data available to the CMCS structure through program modification and data translators. These include data from NIST, NWChem, Ecce, GRI-Mech, Chemkin, and HCT. In addition, the first version of Active Thermochemical Tables (ATcT) has been developed in collaboration with the SciDAC CoG Kit project, with a Web portlet successfully integrated into CMCS. ATcT is a novel scientific application, centered on a distinctively different paradigm of how to obtain reliable thermochemistry based on the thermochemical network approach. The ATcT Web portlet, shown in Figure 2, is available from a project team workspace in the CMCS Portal. The displayed portlet shows a network used to produce an optimized table of thermochemistry data for the methyl peroxy radical (important in low-temperature combustion).

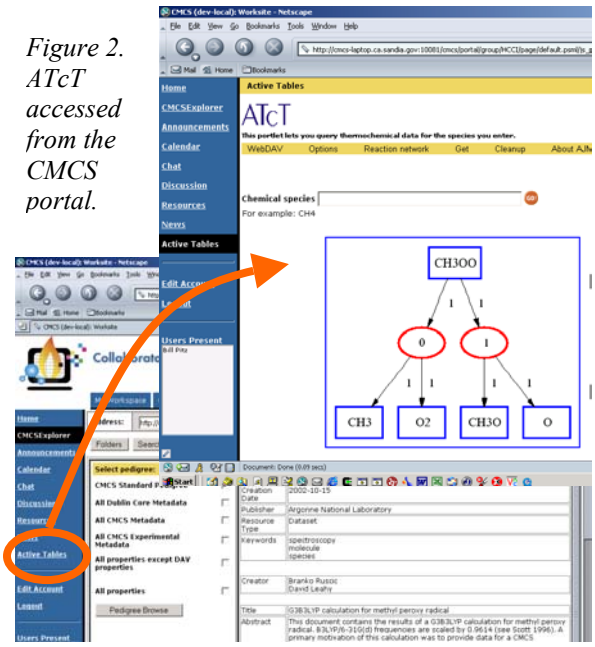

Figure 2. ATcT accessed from the CMCS portal.

In the coming year, CMCS will focus on continued chemical science application integration and pilot community involvement. These include connecting to PrIMe (*Pr*ocess *I*nformatics *M*ode*l*, an international collaboration on chemical mechanisms), to the IUPAC Task on Thermochemistry, to SciDAC quantum chemistry projects, and SciDAC reacting flow projects. (Other collaborations include SAM, CHEF, and CCTTSS & SDM ISICs.)

For further information on this subject contact:
Dr. Larry A. Rahn
Sandia National Laboratories
Phone: 925-294-2091
rahn@sandia.gov
http://cmcs.ca.sandia.gov/